

# Sai Shashank Kudkuli

[github.com/shash992](https://github.com/shash992) | [\(240\) 505-2480](tel:(240)505-2480) | [linkedin.com/in/saishashankk](https://www.linkedin.com/in/saishashankk) | [shashkudkuli@gmail.com](mailto:shashkudkuli@gmail.com) | College Park, MD 20740

## EDUCATION

### University of Maryland

*Master of Science, Information Systems*

August 2024 - December 2025

*College Park, MD, USA*

### PES University

*Bachelor of Technology, Electronics and Communication*

August 2017 - December 2021

*Bengaluru, KA, India*

## TECHNICAL SKILLS

**Core Data & Analytics:** Python, SQL, R, Polars, Spark, Snowflake, HiveQL, Tableau, Power BI, Excel

**AI/ML:** LLM Evaluation, Prompt Engineering, Retrieval-Augmented Generation (RAG), Neo4j, N8N, HuggingFace Transformers, Fine-tuning (Gemma) using LORA optimization, Streamlit

**Cloud & Infrastructure:** AWS, GCP, Azure, Docker, HDFS, Git, CI/CD, CRON, Docker

**Business Tools:** Microsoft Project, Feasibility Analysis, Functional Modeling, Process Documentation

## EXPERIENCE

**R.H. Smith School of Business** | *Research Assistant* | *Snowflake, Transformers* | [aimaps.ai](https://aimaps.ai)

November 2024 - Present

- Optimized Snowflake SQL workflows and engineered batch data pipelines using Polars, reducing end-to-end processing time from 2 hours to 15 minutes across 100M+ job market records; curated and transformed domain-specific datasets for LLM fine-tuning, improving downstream model performance and context relevance.
- Developed ETL systems integrating 5 disparate data sources to support AI Jobs Dashboard for labor market research.
- Designed and implemented LLM evaluation pipelines, benchmarking GPT-4, Claude, Llama-3, and Qwen across 10K+ job-posting dataset, producing model selection insights adopted in academic research and CNN/WSJ-featured reports.

**Oracle Cerner** | *Systems Engineer* | *Linux, Sybase, Active Directory*

September 2022 - December 2023

- Resolved critical Linux application and Sybase database issues for hospital systems, reducing incident tickets from 20 to 5 per week while maintaining 24/7 availability.
- Collaborated with client IT teams to troubleshoot server incidents, achieving 75% improvement mean resolution time.

**Accenture** | *SAP Consultant* | *SAP WM/MM, Service Now*

July 2021 - July 2022

- Analyzed and resolved 500+ SAP issues, maintaining 95% SLA compliance for supply chain operations.
- Created Standard Operating Procedures during ServiceNow migration, reducing process violations by 80%.

## PROJECTS

**Dynamic Surge Pricing using Regression** | *PySpark, Docker, Vertex AI, BigQuery, Pub/Sub* May 2023 – August 2023

- Built a PySpark regression model on Seattle parking dataset to predict availability and apply surge pricing rules, trained on 2022 data and deployed for 2023 live stream.
- Containerized pipeline with Docker, deployed on Vertex AI, automated ingestion via Pub/Sub with results in BigQuery, and visualized surge pricing on Streamlit. Ensured fault-tolerant processing with monitoring via Cloud Logging.

**Graph Search MCP** | *Neo4J, FastMCP* | [github](https://github.com)

June 2025

- Built an AI-ready knowledge graph platform by integrating Neo4j with LLMs via the Model Context Protocol (MCP), enabling scalable natural-language queries over graph data; containerized with Docker for consistency and portability.
- Fine-tuned Gemma 3n using LoRA optimization on relationships, improving relation extraction accuracy.
- Developed and evaluated hybrid graph + LLM retrieval systems, combining symbolic traversal with neural reasoning; achieved a 73% improvement in retrieval accuracy over vector-only baselines.

**Asha IT Recruitment Platform** | *AWS Lambda, DynamoDB, Firebase Auth, Bedrock* | [ashait.com](https://ashait.com)

Jan 2025 - Dec 2025

- Architected an AI-driven recruitment platform using AWS Amplify, Lambda, and DynamoDB for small businesses and international students, integrating Firebase Authentication for secure user management.
- Implemented Agentic AI pipelines in AWS to autonomously match candidates with jobs using LLM-based skill extraction, reasoning over resume-JD embeddings, and multi-agent workflows for ranking and explanation generation.
- Deployed scalable microservices for AI inference via AWS Lambda and API Gateway; optimized query latency by 42% with asynchronous invocation and caching layers.

## LEADERSHIP & RECOGNITION

**Smith Master's IS Association** | *VP Finance*

January 2025 - Present

- Developed mentorship program connecting 100+ students and alumni. [Featured in university article for AI integration.](#)

**Winner:** AI in Business Case Competition: Designed AI integration strategy for small businesses.

November 2024